

# BIG DATA AND BIGGER OPPORTUNITIES

Abdul Baquee  
Muhammad

Islamic Development Bank

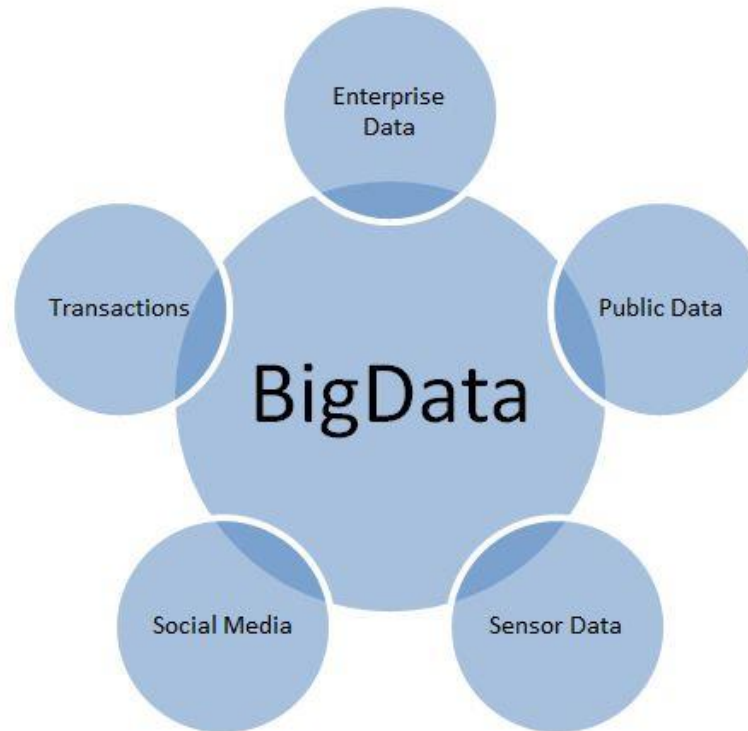
Sixth Session of OIC Statistical  
Commission (OIC-StatCom)

Konya 5-6 Nov. 2016



# HOW BIG IS “BIG DATA”?

3 zettabytes = 3,000,000,000,000,000,000,000 bytes



# THE ARCHITECTURE OF BIG DATA APPLICATIONS

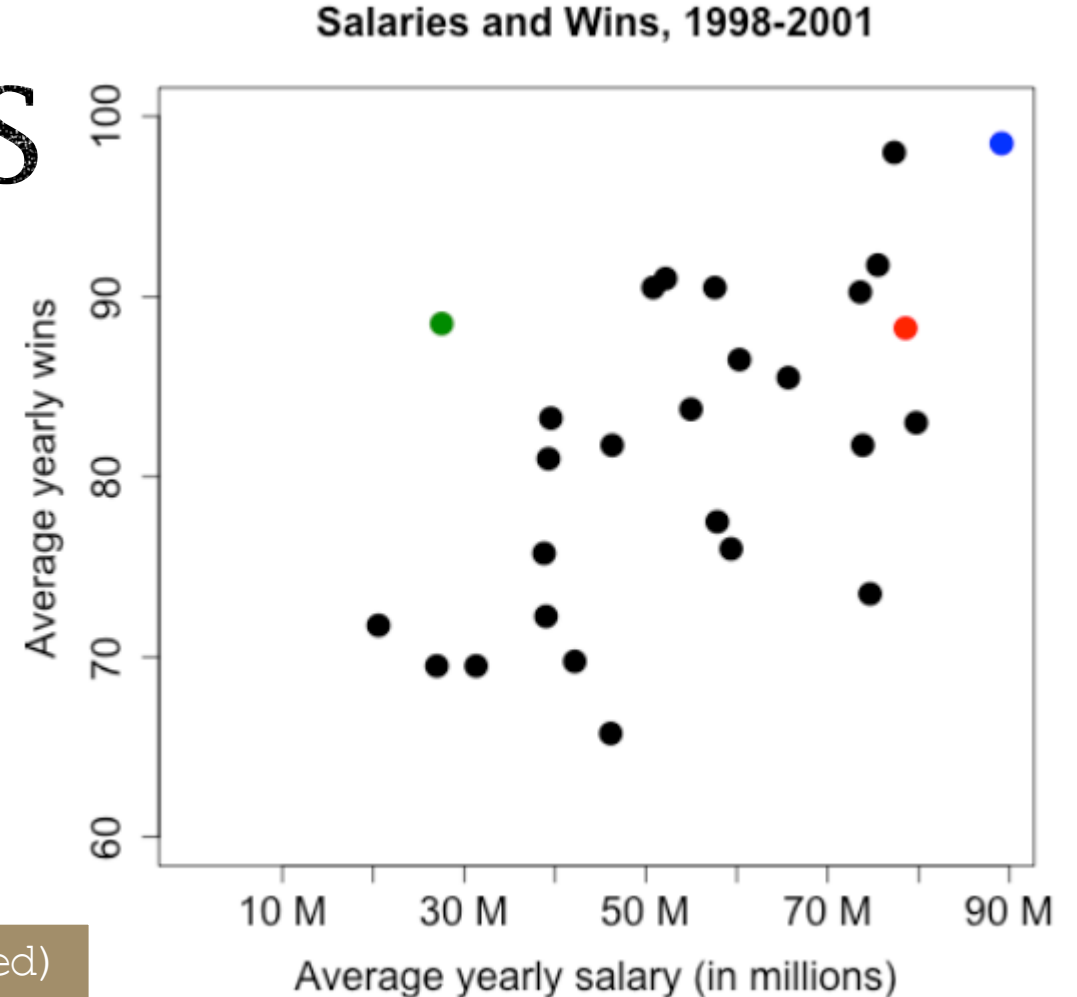


*Source: Machine Learning course @ coursera by Univ. of Washington*

# MONEYBALL & SPORT ANALYTICS

- Quantitative approach to find undervalued players
- Reduce baseball season winning to a math problem
- Developed a linear regression model

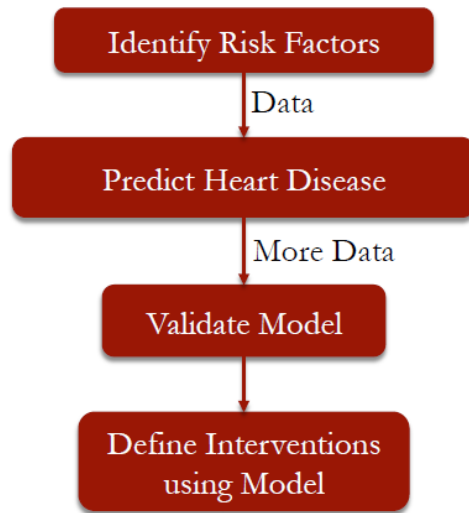
$$\text{Wins} = 80.8814 + 0.1058 (\text{Runs Scored} - \text{Runs Allowed})$$



Source: Analytical Edge course @ EdX by MIT

# PREDICTING HEART ATTACKS

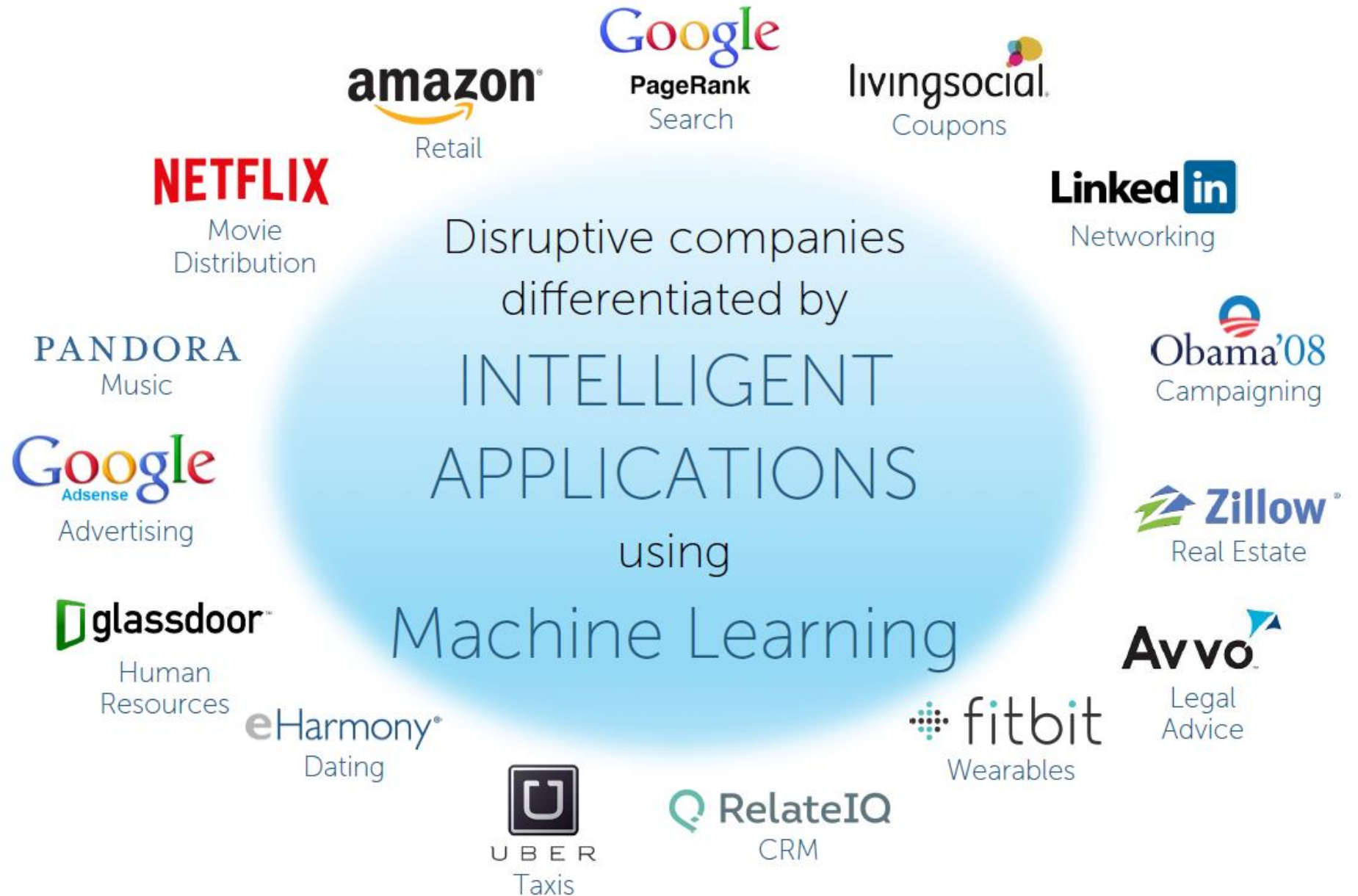
Use logistic regression to predict whether or not a patient would develop heart disease in the next ten years



<http://www.cvriskcalculator.com/>

Age (years)	<input type="text" value="20-79"/>
Gender	<input checked="" type="radio"/> Male <input type="radio"/> Female
Race	<input type="radio"/> African American <input checked="" type="radio"/> Other
Total cholesterol (mg/dL)	<input type="text" value="130-320"/>
HDL cholesterol (mg/dL)	<input type="text" value="20-100"/>
Systolic blood pressure (mmHg)	<input type="text" value="90-200"/>
Diastolic blood pressure (mmHg)	<input type="text" value="30-140"/>
Treated for high blood pressure	<input checked="" type="radio"/> No <input type="radio"/> Yes
Diabetes	<input checked="" type="radio"/> No <input type="radio"/> Yes
Smoker	<input checked="" type="radio"/> No <input type="radio"/> Yes
<input type="button" value="Calculate"/>	

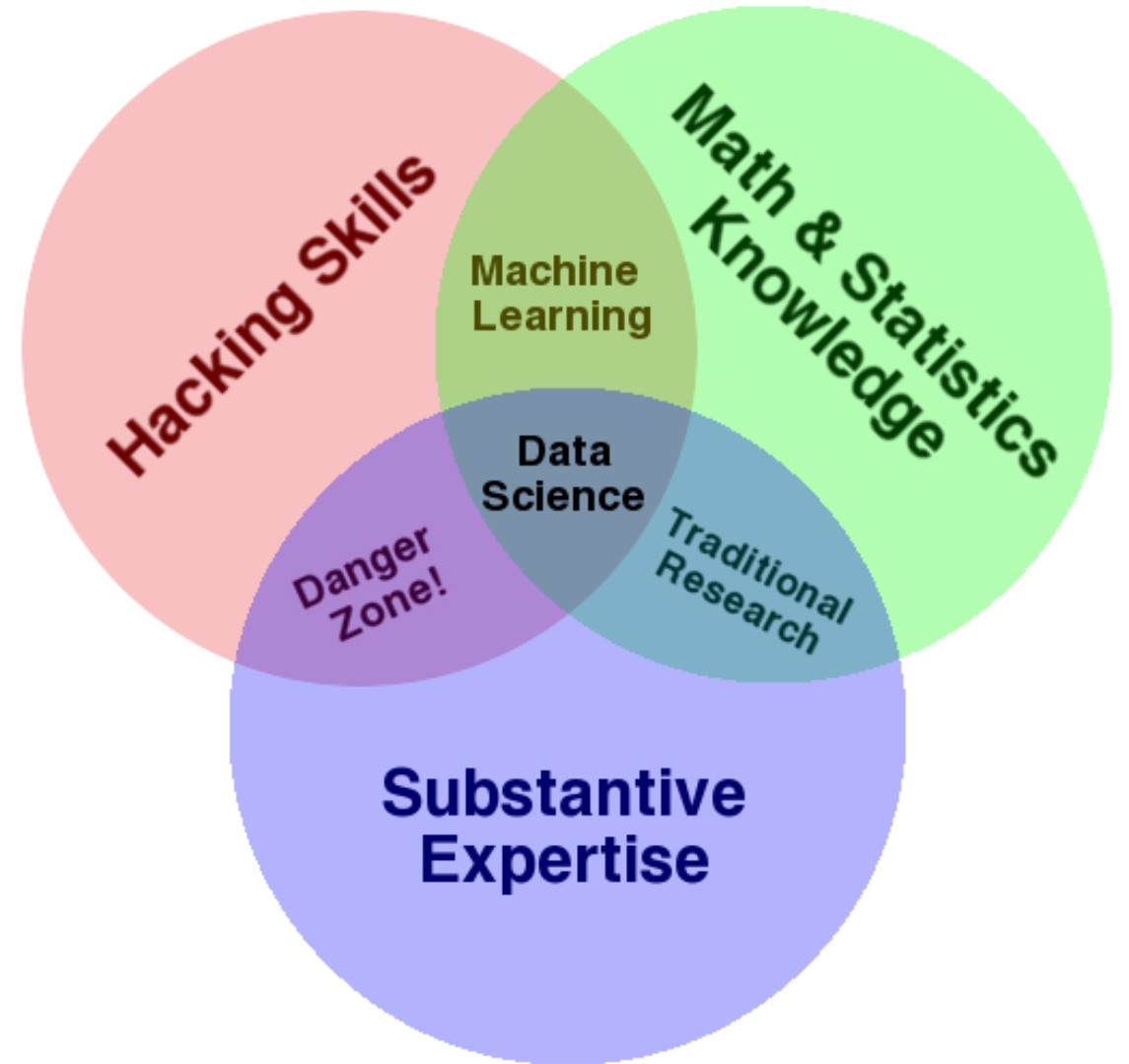
# RIDING THE DATA WAVE @SILICON VALLEY



Source: Machine Learning course @ coursera by Univ. of Washington

# WHAT IS “DATA SCIENCE”

The science of using **data** to build **models** that lead to better **decisions** that add **value** to individuals, to companies, to institutions.



Source: Drew Conway

<http://drewconway.com/zia/2013/3/26/the-data-science-venn-diagram>



# BIG DATA @ DEV





ABOUT US

COMMITMENTS

JOIN

WORKING GROUPS

DATA IN ACTION

NEWS



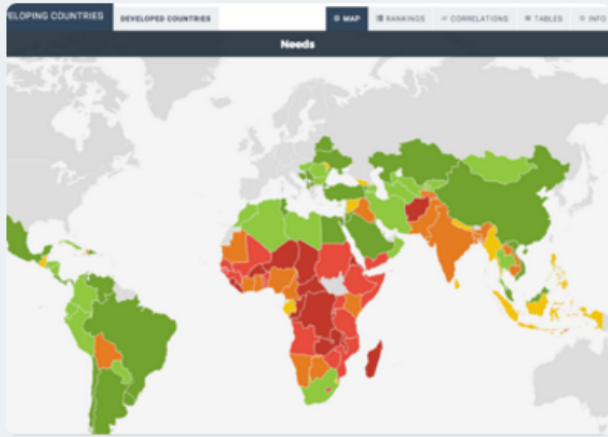
GLOBAL PARTNERSHIP  
FOR SUSTAINABLE DEVELOPMENT DATA

Data4sdg.org

***WE CONVENE  
WE CONNECT  
WE CATALYZE...***

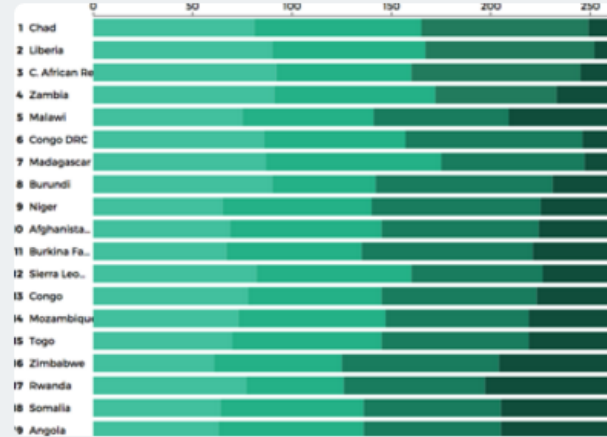
TO HELP END EXTREME POVERTY, FIGHT INEQUALITY AND INJUSTICE, AND COMBAT CLIMATE CHANGE.

# MAP, RANK, AND CORRELATE



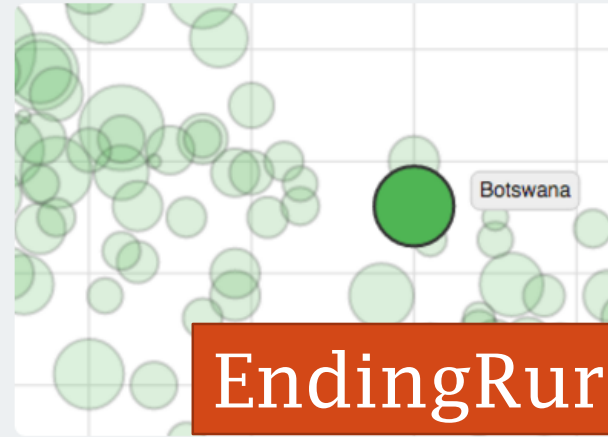
## Map

- ⊕ Where are malnutrition needs highest?
- ⊕ Which countries experience highest vulnerability to consumption shocks?
- ⊕ Which regions have the best policies for FNS?



## Rank

- ⊕ Who ranks most vulnerable on climate change?
- ⊕ Who has the most promising rural investment climates?
- ⊕ Who has the least distorting agricultural and biofuel policies?



## Correlate

- ⊕ How well do developed countries target their FNS aid?
- ⊕ Is there a link between the amount of aid provided and its quality?
- ⊕ Do developing countries with strong FNS policies have lower needs?

ALL DATA TOOLKITS

EndingRuralHunger.Org



# World Bank Big Data Innovation Challenge

Rethinking climate resilience through big data solutions

APPLICATIONS OPEN



## World Bank Big Data Innovation

Academics, entrepreneurs, innovators, businesses, universities and nonprofits have a larger role to play than ever before in addressing global climate issues through the creation and implementation of solutions. The shift to climate-resilient economies can occur only if the millions of decisions which are made across the globe on a daily basis are taking climate change factors and effects into account. We need your help in identifying and developing big data solutions which can help better understand the impacts of climate change, address its connected issues and positively influence decisions.

<https://bigdatainnovationchallenge.org/>

# WORLD BANK BIG DATA CHALLENGE

- Mapping or measuring the risks to farmers' income throughout the food supply chain
- Understanding of the food supply chain, its vulnerabilities and food loss points
- The distribution of value added and income at different stages of the food supply chain
- Understanding, forecasting or raising awareness of constraints on food availability, affordability, and consumption, including local prices and seasonality
- Linking seasonal food production and consumption with nutrition and health, including behavioral aspects, and identification of options for intervention
- Mapping of agricultural production diversity in specific regions, including average crop yields and prices

# Big Data for Social Good Challenge

Powered by IBM + Hadoop

[Home](#)[Participants](#)[Rules](#)[Submissions](#)[Updates](#)[Discussions](#)[IBM Bluemix](#)[Data](#)[FAQ](#)[Videos](#)

IBM Big Data for Social Good Challenge



This hackathon has ended.

Discover more exciting challenges like this one

View the winners

Tell your friends



\$36,000 in prizes

★ Grand Prize

\$20,000 + \$16,000



# IBM Big Data Challenge

**Crime and socioeconomic indicators:** Is there a relationship between the number of crimes in the city of Chicago and socioeconomic indicators (Median income, Poverty, Education, etc)? *Data:* [City of Chicago Crime Data](#), [Median household income by county](#), [Socioeconomic Indicators](#)

**Total 311 complaints locations and median household income:** Can we predict the total 311 complaints per zip code with the median household income in San Francisco? *Data:* [311 complaints in San Francisco](#), [Median household income by county](#)

**Department of homeless services (DHS) and unemployment:** Is there a correlation between the total persons sleeping in shelter each day in NYC and the unemployment rate? *Data:* [DHS daily data](#), [Unemployment Data](#)

**Red light violations and crime rate:** Can we project the corners with high red light violations using crime rate data of each zip code (or county)? *Data:* [Red light violation in Chicago](#), [City of Chicago Crime Data](#)

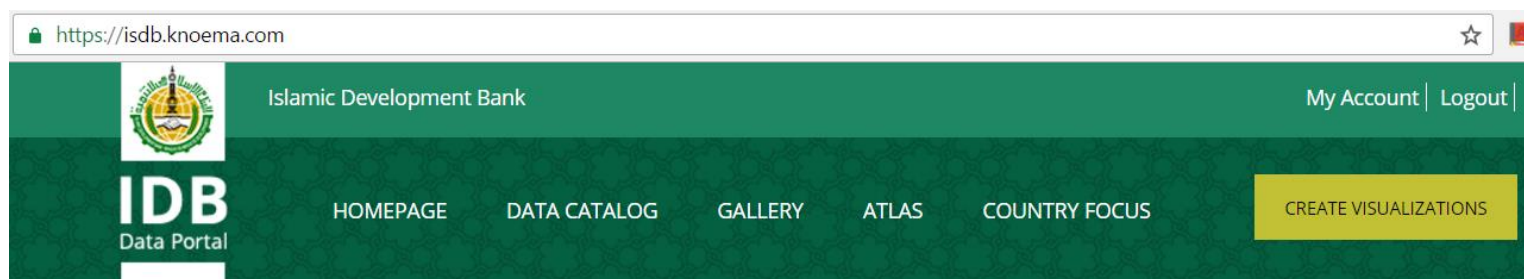
**Pollution, traffic and population of cities:** Can we explain cities that are highly polluted with the population and the traffic data of the city? *Data:* [Pollution](#), [Traffic](#), [Population](#)

**US population and GMOs:** What correlations can you find between US population estimations and the adoption of genetically engineered crops in the U.S ? *Data:* [US Census Population Estimates](#), [ERS GMO crops](#)

**Obesity and food availability:** How does commodity availability and consumption in the United States (fruits, vegetables, meat, grains, etc.) relate to obesity? *Data:* [Obesity Data from NY](#), [USDA ERS Commodity Consumption Data](#)

# VISUALIZATION @ IDB

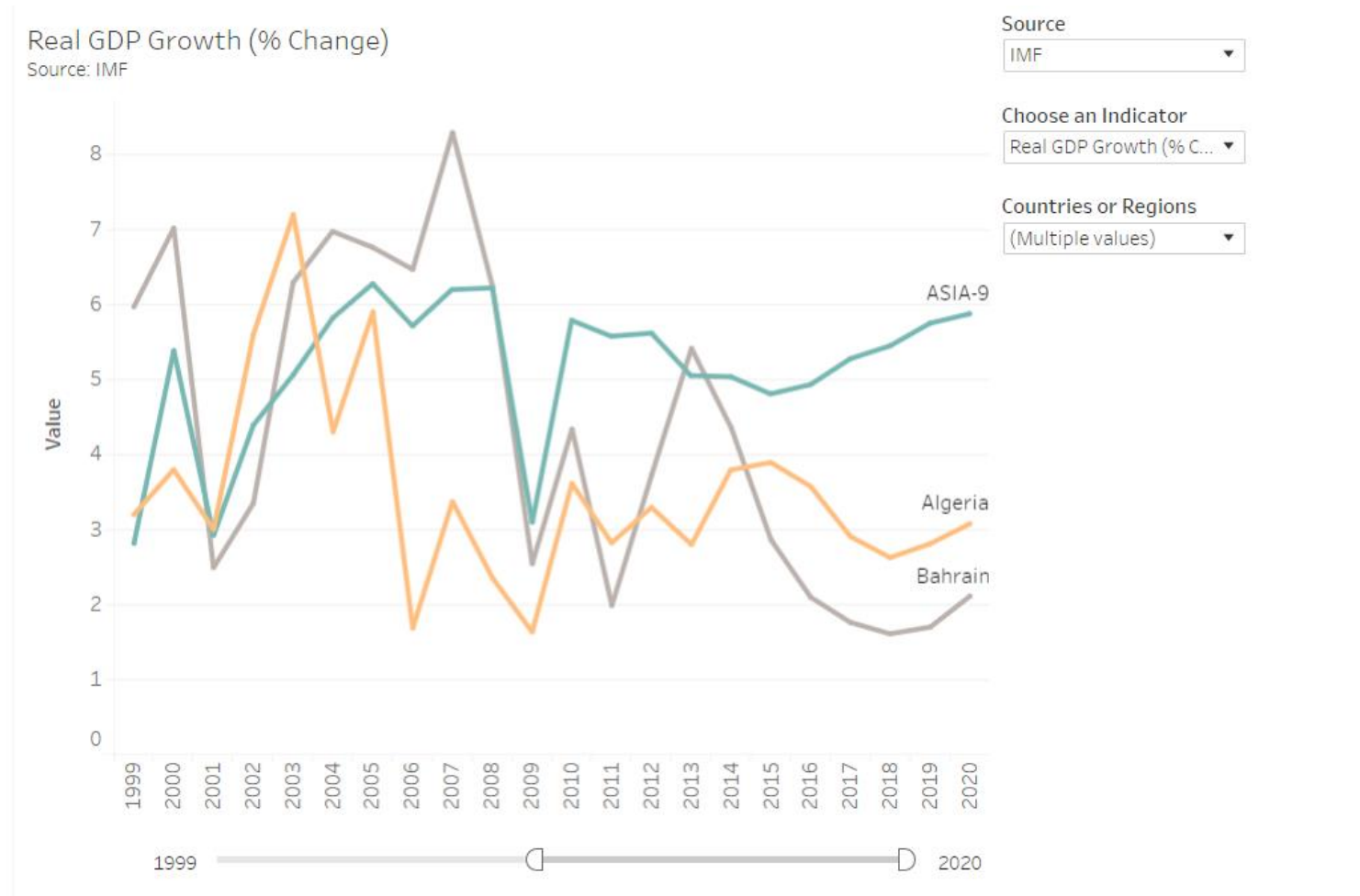
<https://isdb.knoema.com/>



Economic Outlook for IDB Member Countries  
as a Group, 2015

[Browse Data for IDB Member Countries](#)

# CUSTOM VIZ. VIA. TABLEAU



[https://public.tableau.com/views/macro\\_0/dashboard?:embed=y&:display\\_count=yes](https://public.tableau.com/views/macro_0/dashboard?:embed=y&:display_count=yes)



# WAY FORWARD ...

- Build skillsets around 'big data' and 'data science'
- Tons of FREE courses and materials available
  - EdX, Coursera, BigDataUniversity, etc.
- Sponsor interesting 'big data' challenges for social good in our member countries
- Encourage local universities to build curriculums and courses
- Facilitate collection of more and more data